Eigenvalue variance bounds for covariance matrices

Sandrine Dallaporta

Eigenvalue variance bounds for covariance matrices

S. Dallaporta

University of Toulouse and CMLA, ENS Cachan

Abstract. *This work is concerned with finite range bounds on the variance of individual eigenvalues of random covariance matrices, both in the bulk and at the edge of the spectrum. In a preceding paper, the author established analogous results for Wigner matrices [7] and stated the results for covariance matrices. They are proved in the present paper. Relying on the LUE example, which needs to be investigated first, the main bounds are extended to complex covariance matrices by means of the Tao, Vu and Wang Four Moment Theorem and recent localization results by Pillai and Yin. The case of real covariance matrices is obtained from interlacing formulas.*

Random covariance matrices, or Wishart matrices, were introduced by the statistician Wishart in 1928 to model tables of random data in multivariate statistics. The spectral properties of these matrices are indeed useful for example for studying the properties of certain random vectors, elaborating statistical tests and for principal component analysis. Similarly to Wigner matrices, which were introduced by the physicist Wigner in the fifties in order to study infinite-dimensional operators in statistical physics, the asymptotic spectral properties were soon conjectured to be universal in the sense they do not depend on the distribution of the entries (see for example [1] and [19]). Eigenvalues were studied asymptotically both at the global and local regimes, considering for instance the global behavior of the spectrum, the behavior of extreme eigenvalues or the spacings between eigenvalues in the bulk of the spectrum. In the Gaussian case, the eigenvalue joint distribution is explicitly known, allowing for a complete study of the asymptotic spectral properties (see for example [1], [3], [21]). One of the main goals of random matrix theory over the past decades was to extend these results to non-Gaussian covariance matrices.

However, in multivariate statistics, quantitative finite-range results are more useful than asymptotic properties. Furthermore, random covariance matrices have become useful in several other fields, such as compressed sensing (see [31]), wireless communication and quantitative finance (see [3]). In these fields too, quantitative results are of high interest. Several recent developments have thus been concerned with non-asymptotic random matrix theory. See for example some recent surveys and papers on this topic [24], [31] and [30]. In this paper, we investigate in this respect variance bounds on the eigenvalues of families of covariance matrices. In a preceding paper [7], we established similar bounds for Wigner matrices and the results for covariance matrices were stated but not proved. In the present paper, we provide the corresponding proofs. For the sake of completeness and in order to make the present paper readable separately, we reproduce here some parts of the previous one [7].

Random covariance matrices are defined by the following. Let X be a $m \times n$ (real or complex) matrix, with $m \geqslant n$, such that its entries are independent, centered and have variance 1. Then $S_{m,n} = \frac{1}{m} X^* X$ is a covariance matrix. An important example is the case when the entries of X are Gaussian. Then $S_{m,n}$ belongs to the so-called Laguerre Unitary Ensemble (LUE) if the entries of X are complex and to the Laguerre Orthogonal Ensemble (LOE) if they are real. $S_{m,n}$ is Hermitian (or real symmetric) and therefore has n real eigenvalues. As $m \geqslant n$, none of these eigenvalues is trivial. Furthermore, these eigenvalues are nonnegative and will be denoted by $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_n$.

Among universality results, the classical Marchenko-Pastur theorem states that, if $\frac{m}{n} \to \rho \geqslant 1$ when n goes to infinity, the empirical spectral measure $L_{m,n} = \frac{1}{n} \sum_{j=1}^n \delta_{\lambda_j}$ converges almost surely to a deterministic measure, called the Marchenko-Pastur distribution of parameter ρ. This measure is compactly supported and is absolutely continuous with respect to Lebesgue measure, with density

$$d\mu_{MP(\rho)}(x) = \frac{1}{2\pi x} \sqrt{(b_\rho - x)(x - a_\rho)} 1_{[a_\rho, b_\rho]}(x) dx,$$

where $a_\rho = (1 - \sqrt{\rho})^2$ and $b_\rho = (1 + \sqrt{\rho})^2$ (see for example [3]). We denote by $\mu_{m,n}$ the approximate Marchenko-Pastur density

$$\mu_{m,n}(x) = \frac{1}{2\pi x} \sqrt{(x - a_{m,n})(b_{m,n} - x)} 1_{[a_{m,n}, b_{m,n}]}(x),$$

with $a_{m,n} = \left(1 - \sqrt{\frac{m}{n}}\right)^2$ and $b_{m,n} = \left(1 + \sqrt{\frac{m}{n}}\right)^2$. The behavior of individual eigenvalues was more difficult to achieve. At the edge of the spectrum, it was proved by Bai *et al* (see [2], [4] and [5]) under a condition on the fourth moments of the entries that, almost surely,

$$a_{m,n} - \lambda_1 \underset{n \to \infty}{\to} 0 \quad \text{and} \quad \lambda_n - b_{m,n} \underset{n \to \infty}{\to} 0. \tag{1}$$

Once the behavior of eigenvalues at the edge of the spectrum is known, some local information on eigenvalues in the bulk can be deduced from the Marchenko-Pastur theorem. Indeed the Glivenko-Cantelli Theorem gives that almost surely

$$\sup_{x\in\mathbb{R}} |F_{m,n}(x) - G_{m,n}(x)| \underset{n\to\infty}{\to} 0,$$

where $F_{m,n}$ is the distribution function of the empirical spectral distribution $L_{m,n}$ and $G_{m,n}$ is the distribution function of the approximate Marchenko-Pastur law. Combining this with crude bounds on the Marchenko-Pastur density and with (1) leads to the following law of large numbers. For all $\eta > 0$, for all $\eta n \leqslant j \leqslant (1-\eta)n$, i.e. for eigenvalues in the bulk of the spectrum,

$$\lambda_j - \gamma_j^{m,n} \underset{n\to\infty}{\to} 0,$$

almost surely, where the theoretical location $\gamma_j^{m,n} \in [a_{m,n}, b_{m,n}]$ of the j-th eigenvalue λ_j is defined by

$$\frac{j}{n} = \int_{a_{m,n}}^{\gamma_j^{m,n}} \mu_{m,n}(x)\,dx.$$

At the fluctuation level, the behavior of individual eigenvalues depends heavily on their location in the spectrum and on the value of the parameter ρ, at least for the smallest eigenvalues. Indeed, when $\rho > 1$, the left-side of the limiting support a_ρ is positive. As a consequence, eigenvalues, and in particular smallest eigenvalues, can be less than a_ρ, which is therefore called a soft edge. On the contrary, when $\rho = 1$, $a_\rho = 0$ and no eigenvalue can be less than a_ρ. In this case, the left-side is called a hard edge. Even the behavior of the Marchenko-Pastur density is different at the lower edge in these two cases. Indeed, when $\rho > 1$, the Marchenko-Pastur density function is bounded whereas it goes to ∞ when $x \to 0$ if $\rho = 1$. Therefore, the behavior of the smallest eigenvalue is expected to be different according to ρ. Indeed, on the one hand, when $m = n$ (which implies $\rho = 1$), Edelman proved that, for LUE matrices,

$$n^2\lambda_1 \underset{n\to\infty}{\overset{(d)}{\to}} \mathcal{E}(1),$$

where $\mathcal{E}(1)$ is an exponential random variable with parameter 1. A similar result is available for LOE matrices, see [8] for more details. This theorem was later extended to more general covariance matrices by Tao and Vu in [27]. On the other hand, when $\rho > 1$, Borodin and Forrester proved that, for LUE matrices,

$$n^{2/3}\frac{a_{m,n} - \lambda_1}{a_{m,n}^{2/3}\left(\frac{m}{n}\right)^{-1/6}} \underset{n\to\infty}{\overset{(d)}{\to}} F_2,$$

where F_2 is the so-called Tracy-Widom law (see [6]). A similar result holds for LOE matrices. These theorems were later extended to some non-Gaussian covariance matrices by Feldheim and Sodin in [10] and then to large families of covariance matrices by Wang (see [33]). On the contrary, the behavior of the largest eigenvalue relies much less on the value of the parameter ρ. Indeed Johansson (see [16]) proved that, for LUE matrices,

$$n^{2/3} \frac{\lambda_n - b_{m,n}}{b_{m,n}^{2/3} \left(\frac{m}{n}\right)^{-1/6}} \xrightarrow[n\to\infty]{(d)} F_2.$$

Johnstone proved a similar result for LOE matrices (see [17]). Soshnikov and Péché extended these theorems to more general covariance matrices in [25] and [22]. They were then extended to large families of non-Gaussian covariance matrices by Wang in [33]. From these central limit theorems, the variances of the smallest (when $\rho > 1$) and largest eigenvalues are guessed to be of the order of $n^{-4/3}$.

In the bulk of the spectrum, i.e. for all eigenvalues λ_j such that $\eta n \leqslant j \leqslant (1-\eta)n$ for a fixed $\eta > 0$, Su proved in [26] that

$$\mu_{m,n}(\gamma_j^{m,n}) \frac{\lambda_j - \gamma_j^{m,n}}{\sqrt{\frac{1}{2\pi^2} \frac{\log n}{n^2}}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1),$$

in distribution. As for the largest eigenvalue, the value of parameter ρ does not change significantly the behavior of eigenvalues in the bulk. This Central Limit Theorem was extended to families of non-Gaussian matrices by Tao and Vu in [28]. The variances of eigenvalues in the bulk are then guessed to be of the order of $\frac{\log n}{n^2}$. Su proved in [26] a similar Central Limit Theorem for right-side intermediate eigenvalues, which means eigenvalues λ_j with $\frac{j}{n} \to 1$ and $n - j \to \infty$ when n goes to infinity. From this theorem, the variance of such eigenvalues is guessed to be of the order of $\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}$. This theorem was later extended to non-Gaussian covariance matrices by Wang in [33]. It seems that a similar result holds for left-side intermediate eigenvalues when $\rho > 1$ but Su did not carry out the computations in this case.

The aim of this paper is to provide sharp non-asymptotic bounds for the variance of individual eigenvalues of covariance matrices. For simplicity, we basically assume that $\rho > 1$. More precisely, we assume that $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2$ (where A_1 and A_2 are fixed constants). When $m = n$ (therefore $\rho = 1$), it is possible to show that the following results in the bulk and on the right-side of the spectrum are true. It will be specified in the corresponding sections. Assume furthermore that $S_{m,n}$ is a complex covariance matrix (respectively real) whose entries have an exponential decay and have the same first four moments as those of a LUE (respectively LOE) matrix. This condition is called condition $(C0)$ and will be detailed in Section 2. The main results of this paper are the following theorems.

Theorem 1 (in the bulk of the spectrum)**.** *For all $0 < \eta \leqslant \frac{1}{2}$, there exists a constant $C > 0$ depending only on η, A_1 and A_2 such that, for all $\eta n \leqslant j \leqslant (1-\eta)n$,*

$$\mathrm{Var}(\lambda_j) \leqslant C\frac{\log n}{n^2}. \tag{2}$$

Theorem 2 (between the bulk and the edge of the spectrum)**.** *There exists a constant $\kappa > 0$ (depending on A_1 and A_2) such that the following holds. For all $K > \kappa$, for all $0 < \eta \leqslant \frac{1}{2}$, there exists a constant $C > 0$ (depending on K, η, A_1 and A_2) such that for all covariance matrix $S_{m,n}$, for all $(1-\eta)n \leqslant j \leqslant n-K\log n$,*

$$\mathrm{Var}(\lambda_j) \leqslant C\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}. \tag{3}$$

Theorem 3 (at the edge of the spectrum)**.** *There exists a constant $C > 0$ depending only on A_1 and A_2 such that,*

$$\mathrm{Var}(\lambda_n) \leqslant Cn^{-4/3}. \tag{4}$$

It should be mentioned that Theorem 2 (respectively Theorem 3) probably holds for left-hand side intermediate eigenvalues (respectively the smallest eigenvalue λ_1), when $\rho > 1$. We refer to Section 1.2 for more details on that topic. On the contrary, when $\rho = 1$, the behavior of eigenvalues on the left-side of the spectrum is probably very different and much more difficult to study.

The first two theorems do not seem to be known even for LUE matrices. The first step is then to establish these results for such matrices. The proof relies on the fact that the eigenvalues of a LUE matrix form a determinantal process and therefore that the eigenvalue counting function has the same distribution as a sum of independent Bernoulli variables [15]. Using a standard concentration inequality for Bernoulli variables, it is then possible to establish a deviation inequality for individual eigenvalues. A simple integration leads to the desired bounds on the variances. On the contrary, Theorem 3 on the largest eigenvalue λ_n of LUE matrices has been known for some time, at least for the largest eigenvalue λ_n (see [18]). From these results for the LUE, Theorems 1, 2 and 3 are then extended to large families of non-Gaussian covariance matrices by means of localization properties by Pillai and Yin (see [23]) and the Four Moment Theorem by Tao-Vu and Wang (see [28] and [33]). While the localization properties almost yield the correct order on the variance, the Four Moment Theorem is used to reach the optimal bound via a comparison with LUE matrices. Theorems 1, 2 and 3 are established first in the complex case. The real case is then achieved by means of interlacing formulas. Note that similar inequalities hold for higher moments of the eigenvalues. The proofs are exactly the same.

As a corollary of the preceding results on the variances and provided Theorem 2 holds also for left-hand side intermediate eigenvalues, a bound on the rate of convergence of the empirical spectral distribution $L_{m,n}$ towards the Marchenko-Pastur distribution can be achieved. It can be written in terms of the 2-Wasserstein distance between the approximate Marchenko-Pastur distribution $\mu_{m,n}$ and $L_{m,n}$, defined by the following. For μ and ν two probability measures on $\mathbb{R}$,

$$W_2(\mu,\nu) = \inf \left(\int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2},$$

where the infimum is taken over all probability measure π on $\mathbb{R}^2$ such that its first marginal is μ and its second marginal is ν. Note that the rate of convergence of this empirical distribution has also been investigated in terms of the Kolmogorov distance between $L_{m,n}$ and $\mu_{m,n}$ (see for example [12] and [13]). This distance is defined by

$$d_K(L_{m,n},\mu_{m,n}) = \sup_{x\in\mathbb{R}} \left| \tfrac{1}{n}\mathcal{N}_x - G_{m,n}(x) \right|,$$

where $\mathcal{N}_x$ is the eigenvalue counting function, i.e. $\mathcal{N}_x$ is the number of eigenvalues in $(-\infty, x]$, and $G_{m,n}$ is the distribution function of the approximate Marchenko-Pastur distribution. Götze and Tikhomirov recently showed that, with high probability,

$$d_K(L_{m,n},\mu_{m,n}) \leqslant \frac{(\log n)^c}{n}$$

for some universal constant $c > 0$ (see [13]). The rate of convergence in terms of the 1-Wasserstein distance W_1, also called the Kantorovich-Rubinstein distance, was studied by Guionnet and Zeitouni in [14], who proved that $\mathrm{E}[W_1(L_{m,n}, \mathrm{E}[L_{m,n}])]$ is bounded by $Cn^{-2/5}$. The following statement is concerned with the expectation of $W_2(L_{m,n},\mu_{m,n})$.

Corollary 4. *Let* $1 < A_1 < A_2$. *Then there exists a constant* $C > 0$ *depending only on* A_1 *and* A_2 *such that, for all* m *and* n *such that* $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2$,

$$\mathrm{E}\Big[W_2^2(L_{m,n},\mu_{m,n})\Big] \leqslant C\frac{\log n}{n^2}. \tag{5}$$

The proof of this corollary relies on the fact that $\mathrm{E}[W_2^2(L_{m,n},\mu_{m,n})]$ is bounded above, up to a constant, by the sum of the expectations $\mathrm{E}[(\lambda_j - \gamma_j^{m,n})^2]$. The previously established bounds then easily yield the result, provided Theorem 2 holds for left-hand side intermediate eigenvalues.

Turning now to the content of this paper, Section 1 describes Theorems 1, 2 and 3 in the LUE case. Section 2 starts with the Localization Theorem of Pillai and Yin (see [23]) and the Four Moment Theorem of Tao-Vu and Wang (see [28] and [33]). Theorems 1, 2 and 3 are then established for families of covariance matrices. Section 3 is devoted to real matrices. Section 4 deals with Corollary 4 and the rate of convergence of $L_{m,n}$ towards $\mu_{MP(\rho)}$ in terms of 2-Wasserstein distance.

Throughout this paper, C and c will denote positive constants, which depend on the indicated parameters and whose values may change from one line to another.

1 Deviation inequalities and variance bounds for LUE matrices

This section is concerned with Gaussian covariance matrices. The results and techniques used here heavily rely on the Gaussian structure, in particular on the determinantal properties of the eigenvalues. As a consequence of this determinantal structure, the eigenvalue counting function is known to have the same distribution as a sum of independent Bernoulli variables (see [15], [1]). Its mean and variance were computed by Su (see [26]). Deviation inequalities can therefore be established for individual eigenvalues, leading to the announced bounds on the variance. All the proofs are written in the case when $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2$. Assuming $m = n$, if the results still hold, the proofs are very similar and are therefore not reproduced.

1.1 Inside the bulk of the spectrum

The aim of this section is to prove the following theorem for eigenvalues in the bulk, i.e. for λ_j with $\eta n \leqslant j \leqslant (1-\eta)n$.

Theorem 5. *Let $1 < A_1 < A_2$. Let $S_{m,n}$ be a LUE matrix. For any $0 < \eta \leqslant \frac{1}{2}$, there exists a constant $C > 0$ depending only on η, A_1 and A_2 such that for all $A_1 \leqslant \frac{m}{n} \leqslant A_2$ and all $\eta n \leqslant j \leqslant (1-\eta)n$,*

$$\mathrm{E}\Big[|\lambda_j - \gamma_j^{m,n}|^2\Big] \leqslant C\frac{\log n}{n^2}. \tag{6}$$

In particular,

$$\mathrm{Var}(\lambda_j) \leqslant C\frac{\log n}{n^2}. \tag{7}$$

The proof of this theorem relies on the properties of the eigenvalue counting function, denoted by $\mathcal{N}_t = \sum_{i=1}^n 1_{\lambda_i \leqslant t}$ for every $t \in \mathbb{R}$. As announced, $\mathcal{N}_t$ has the same distribution as a sum of independent Bernoulli variables [15]. Consequently, sharp deviation inequalities are available for $\mathcal{N}_t$. Applying Bernstein's inequality leads to

$$\mathrm{P}\Big(\big|\mathcal{N}_t - \mathrm{E}[\mathcal{N}_t]\big| \geqslant u\Big) \leqslant 2\exp\Big(-\frac{u^2}{2\sigma_t^2 + u}\Big), \tag{8}$$

where σ_t^2 is the variance of $\mathcal{N}_t$ (see for example [29]). Götze and Tikhomirov proved in [12] that, as soon as $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2$, there exists a positive constant C_1 depending only on A_1 and A_2 such that

$$\sup_{t\in\mathbb{R}} \left| \mathrm{E}[\mathcal{N}_t] - n\int_{-\infty}^{t} \mu_{m,n}(x)\,dx \right| \leqslant C_1, \tag{9}$$

for every LUE matrix $S_{m,n}$. For simplicity, we denote $\int_{-\infty}^{t} \mu_{m,n}(x)\,dx$ by μ_t. Together with (8), for every $u \geqslant 0$,

$$\mathrm{P}\Big(|\mathcal{N}_t - n\mu_t| \geqslant u + C_1 \Big) \leqslant 2\exp\Big(-\frac{u^2}{2\sigma_t^2 + u} \Big). \tag{10}$$

Among Su's results (see [26]), for every $\delta > 0$, there exists $c_\delta > 0$ such that

$$\sup_{t\in I_\delta} \sigma_t^2 \leqslant c_\delta \log n, \tag{11}$$

where $I_\delta = [a_{m,n} + \delta, b_{m,n} - \delta]$. Combining (10) and (11), deviation inequalities for individual eigenvalues in the bulk are then available, as stated in the following proposition.

Proposition 6. *Assume that $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2$. Let $\eta > 0$. Then there exist positive constants C, c, c' and δ such that the following holds. For any LUE matrix $S_{m,n}$, for all $\eta n \leqslant j \leqslant (1-\eta)n$, for all $c \leqslant u \leqslant c'n$,*

$$\mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| \geqslant \frac{u}{n} \Big) \leqslant 4\exp\Big(-\frac{C^2u^2}{2c_\delta \log n + Cu} \Big). \tag{12}$$

The constants C, c' and δ depend only on η and A_2, whereas the constant c depends only on η, A_1 and A_2.

Note that this proposition still holds if $m = n$, as Götze and Tikhomirov proved in [12] that (9) holds in that case.

Proof. Let $\eta > 0$ and $u \geqslant 0$. Assume first that $\frac{n}{2} \leqslant j \leqslant (1-\eta)n$. Start with estimating the probability that λ_j is greater than $\gamma_j^{m,n} + \frac{u}{n}$.

$$\begin{aligned}
\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n} + \frac{u}{n} \Big) &= \mathrm{P}\Big(\mathcal{N}_{\gamma_j^{m,n}+\frac{u}{n}} < j \Big) \\
&= \mathrm{P}\Big(n\mu_{\gamma_j^{m,n}+\frac{u}{n}} - \mathcal{N}_{\gamma_j^{m,n}+\frac{u}{n}} > n\mu_{\gamma_j^{m,n}+\frac{u}{n}} - j \Big) \\
&= \mathrm{P}\Big(n\mu_{\gamma_j^{m,n}+\frac{u}{n}} - \mathcal{N}_{\gamma_j^{m,n}+\frac{u}{n}} > n(\mu_{\gamma_j^{m,n}+\frac{u}{n}} - \mu_{\gamma_j^{m,n}}) \Big) \\
&\leqslant \mathrm{P}\Big(|\mathcal{N}_{\gamma_j^{m,n}+\frac{u}{n}} - n\mu_{\gamma_j^{m,n}+\frac{u}{n}}| > n(\mu_{\gamma_j^{m,n}+\frac{u}{n}} - \mu_{\gamma_j^{m,n}}) \Big),
\end{aligned}$$

where it has been used that $\mu_{\gamma_j^{m,n}} = \frac{j}{n}$. In order to use (10), a lower bound on $n(\mu_{\gamma_j^{m,n}+\frac{u}{n}} - \mu_{\gamma_j^{m,n}})$ is needed.

$$\begin{aligned}
\mu_{\gamma_j^{m,n}+\frac{u}{n}} - \mu_{\gamma_j^{m,n}} &= \int_{\gamma_j^{m,n}}^{\gamma_j^{m,n}+\frac{u}{n}} \frac{1}{2\pi x}\sqrt{(b_{m,n}-x)(x-a_{m,n})}\,dx \\
&\geqslant \frac{\sqrt{\gamma_j^{m,n}-a_{m,n}}}{2\pi b_{m,n}} \int_{\gamma_j^{m,n}}^{\gamma_j^{m,n}+\frac{u}{n}} \sqrt{b_{m,n}-x}\,dx \\
&\geqslant \frac{\sqrt{\gamma_j^{m,n}-a_{m,n}}}{3\pi b_{m,n}} (b_{m,n}-\gamma_j^{m,n})^{3/2}\Big(1-\Big(1-\frac{u/n}{b_{m,n}-\gamma_j^{m,n}}\Big)^{3/2}\Big) \\
&\geqslant \frac{\sqrt{\gamma_j^{m,n}-a_{m,n}}}{3\pi b_{m,n}} (b_{m,n}-\gamma_j^{m,n})^{1/2}\frac{u}{n},
\end{aligned}$$

if $\gamma_j^{m,n} + \frac{u}{n} \leqslant b_{m,n}$. Furthermore, by definition of $\gamma_j^{m,n}$,

$$\begin{aligned}
1-\frac{j}{n} &= \int_{\gamma_j^{m,n}}^{b_{m,n}} \frac{1}{2\pi x}\sqrt{(x-a_{m,n})(b_{m,n}-x)}\,dx \\
&\leqslant \frac{\sqrt{b_{m,n}-a_{m,n}}}{2\pi\gamma_j^{m,n}} \int_{\gamma_j^{m,n}}^{b_{m,n}} \sqrt{b_{m,n}-x}\,dx \\
&\leqslant \frac{\sqrt{b_{m,n}-a_{m,n}}}{3\pi\gamma_j^{m,n}} \Big(b_{m,n}-\gamma_j^{m,n}\Big)^{3/2}.
\end{aligned}$$

Then

$$b_{m,n}-\gamma_j^{m,n} \geqslant \Big(\frac{3\pi}{\sqrt{b_{m,n}-a_{m,n}}}\gamma_j^{m,n}\Big(1-\frac{j}{n}\Big)\Big)^{2/3}. \tag{13}$$

Moreover

$$\begin{aligned}
\frac{j}{n} &= \int_{a_{m,n}}^{\gamma_j^{m,n}} \frac{1}{2\pi x}\sqrt{(x-a_{m,n})(b_{m,n}-x)}\,dx \\
&\leqslant \frac{\sqrt{b_{m,n}-a_{m,n}}}{2\pi} \int_{a_{m,n}}^{\gamma_j^{m,n}} \frac{1}{x}\sqrt{x-a_{m,n}}\,dx \\
&\leqslant \frac{\sqrt{b_{m,n}-a_{m,n}}}{2\pi} \int_0^{\sqrt{\gamma_j^{m,n}-a_{m,n}}} \frac{2v^2}{v^2+a_{m,n}}\,dv,
\end{aligned}$$

with change of variables $x = v^2 + a_{m,n}$. Therefore,

$$\frac{j}{n} \leqslant \frac{\sqrt{b_{m,n}-a_{m,n}}}{\pi}\sqrt{\gamma_j^{m,n}-a_{m,n}}.$$

Then

$$\sqrt{\gamma_j^{m,n}} \geqslant \sqrt{\gamma_j^{m,n} - a_{m,n}} \geqslant \frac{\pi}{\sqrt{b_{m,n} - a_{m,n}}} \frac{j}{n} \geqslant \frac{\pi}{2\sqrt{b_{m,n} - a_{m,n}}}, \tag{14}$$

as $j \geqslant \frac{n}{2}$. Therefore, as $1 - \frac{j}{n} \geqslant \eta$,

$$b_{m,n} - \gamma_j^{m,n} \geqslant \left(\tfrac{3}{4}\right)^{2/3} \frac{\pi^2}{b_{m,n} - a_{m,n}} \eta^{2/3}. \tag{15}$$

As a consequence, a lower bound on $\mu_{\gamma_j^{m,n} + \frac{u}{n}} - \mu_{\gamma_j^{m,n}}$ is achieved.

$$\mu_{\gamma_j^{m,n} + \frac{u}{n}} - \mu_{\gamma_j^{m,n}} \geqslant \frac{C}{b_{m,n}(b_{m,n} - a_{m,n})} \eta^{1/3} \frac{u}{n},$$

where $C > 0$ is a universal constant. As $\frac{m}{n} \leqslant A_2$,

$$\mu_{\gamma_j^{m,n} + \frac{u}{n}} - \mu_{\gamma_j^{m,n}} \geqslant \frac{C}{4\sqrt{A_2}(1 + \sqrt{A_2})^2} \eta^{1/3} \frac{u}{n} = C(A_2, \eta) \frac{u}{n}.$$

Then

$$\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n} + \frac{u}{n}\Big) \leqslant \mathrm{P}\Big(|\mathcal{N}_{\gamma_j^{m,n} + \frac{u}{n}} - n\mu_{\gamma_j^{m,n} + \frac{u}{n}}| > C(A_2, \eta) u\Big).$$

This is true for all $u \leqslant n(b_{m,n} - \gamma_j^{m,n})$. From (15), this will be true when $u \leqslant c'n$ where $c' > 0$ depends only on A_2 and η. If $u \geqslant c = \frac{2C_1}{C(A_2, \eta)}$, then

$$\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n} + \frac{u}{n}\Big) \leqslant \mathrm{P}\Big(|\mathcal{N}_{\gamma_j^{m,n} + \frac{u}{n}} - n\mu_{\gamma_j^{m,n} + \frac{u}{n}}| > \tfrac{1}{2} C(A_2, \eta) u + C_1\Big).$$

Consequently, from (10), we get

$$\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n} + \frac{u}{n}\Big) \leqslant 2 \exp\Big(-\frac{u^2}{2\sigma^2_{\gamma_j^{m,n} + \frac{u}{n}} + u}\Big).$$

For $u \leqslant c'n$ (with a maybe smaller $c' > 0$ depending only on η and A_2), there exists $\delta > 0$ depending on η and A_2 such that $\gamma_j^{m,n} + \frac{u}{n} \in I_\delta$. Consequently, from (11), $\sigma^2_{\gamma_j^{m,n} + \frac{u}{n}} \leqslant c_\delta \log n$. Then, for $c \leqslant u \leqslant c'n$,

$$\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n} + \frac{u}{n}\Big) \leqslant 2 \exp\Big(-\frac{u^2}{2c_\delta \log n + u}\Big).$$

Repeating the argument leads to the same bound on $\mathrm{P}\Big(\lambda_j < \gamma_j^{m,n} - \frac{u}{n}\Big)$. Therefore,

$$\mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| \geqslant \frac{u}{n}\Big) \leqslant 4 \exp\Big(-\frac{C^2 u^2}{2c_\delta \log n + Cu}\Big).$$

The case when $j \leqslant \frac{n}{2}$ is treated similarly. The proposition is thus established.

□

We turn now to the proof of Theorem 5.

Proof of Theorem 5. Note first that, for every i,

$$\mathrm{E}[\lambda_i^4] \leqslant \sum_{j=1}^{n} \mathrm{E}[\lambda_j^4] = \mathrm{E}[\mathrm{Tr}(S_{m,n}^4)].$$

From Hölder inequality,

$$\mathrm{E}[\mathrm{Tr}(S_{m,n}^4)] \leqslant \frac{1}{n^4} \sum_{\substack{j_1,\dots,j_4 \in [\![1,m]\!] \\ i_1,\dots,i_4 [\![1,n]\!]}} \Big(\mathrm{E}[|X_{i_1,j_1}|^8]\Big)^{1/8} \dots \Big(\mathrm{E}[|X_{i_4,j_1}|^8]\Big)^{1/8}. \qquad (16)$$

As $S_{m,n}$ is from the LUE, the 8^{th} moment of its entries is $\mathrm{E}[|X_{i,j}|^8] = 105$. Then

$$\mathrm{E}[\mathrm{Tr}(S_{m,n}^4)] \leqslant 105 m^4 \leqslant 105 A_2^4 n^4.$$

Consequently, for all $n \geqslant 1$, for all $1 \leqslant i \leqslant n$,

$$\mathrm{E}[\lambda_i^4] \leqslant 105 A_2^4 n^4. \qquad (17)$$

Consider constants C, c, c' and δ given by Proposition 6. Choose next $M > 0$ large enough such that $\frac{C^2M^2}{2c_\delta + CM} > 8$. M depends only on η and A_2. Setting $Z = n|\lambda_j - \gamma_j^{m,n}|$,

$$\begin{aligned}
\mathrm{E}[Z^2] &= \int_0^\infty \mathrm{P}(Z \geqslant v) 2v \, dv \\
&= \int_0^c \mathrm{P}(Z \geqslant v) 2v \, dv + \int_c^{M \log n} \mathrm{P}(Z \geqslant v) 2v \, dv + \int_{M \log n}^\infty \mathrm{P}(Z \geqslant v) 2v \, dv \\
&\leqslant c^2 + I_1 + I_2.
\end{aligned}$$

The two latter integrals are handled in different ways. The first one I_1 is bounded using (12) while I_2 is controlled using the Cauchy-Schwarz inequality and (17). Starting thus with I_2,

$$\begin{aligned}
I_2 &= \int_{M \log n}^{+\infty} \mathrm{P}(Z \geqslant v) 2v \, dv \\
&\leqslant \mathrm{E}\Big[Z^2 1_{Z \geqslant M \log n}\Big] \\
&\leqslant \sqrt{\mathrm{E}\Big[Z^4\Big]} \sqrt{\mathrm{P}\Big(Z \geqslant M \log n\Big)} \\
&\leqslant A n^4 \sqrt{\mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| \geqslant \frac{M \log n}{n}\Big)} \\
&\leqslant 2 A n^4 \exp\Big(-\frac{1}{2} \frac{C^2 M^2}{2c_\delta + CM} \log n\Big) \\
&\leqslant 2 A \exp\bigg(\frac{1}{2}\Big(8 - \frac{C^2 M^2}{2c_\delta + CM}\Big) \log n\bigg),
\end{aligned}$$

where $A > 0$ is a numerical constant. As $\exp\Big(\frac{1}{2}\Big(8 - \frac{C^2M^2}{2c_\delta + CM}\Big)\log n\Big) \underset{n\to\infty}{\to} 0$, there exists a constant $C > 0$ (depending only on η and A_2) such that

$$I_2 \leqslant C.$$

Turning to I_1, recall that Proposition 6 gives, for $c \leqslant v \leqslant c'n$,

$$P(Z \geqslant v) = \mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| \geqslant \frac{v}{n}\Big) \leqslant 4\exp\Big(-\frac{C^2v^2}{2c_\delta \log n + Cv}\Big).$$

Hence in the range $v \leqslant M \log n$,

$$P(Z \geqslant v) \leqslant 4\exp\Big(-\frac{B}{\log n}v^2\Big),$$

where $B = B(A_2, \eta) = \frac{C^2}{2c_\delta + CM}$. As a consequence,

$$I_1 \leqslant 4\int_c^{M\log n} \exp\Big(-\frac{B}{\log n}v^2\Big)2v\,dv \leqslant \frac{4\log n}{B}\int_0^\infty e^{-v^2}2v\,dv.$$

There exists thus a constant $C > 0$ (depending only on η and A_2) such that

$$I_1 \leqslant C\log n.$$

Summarizing the previous steps, $\mathrm{E}\Big[Z^2\Big] \leqslant C\log n$. Therefore

$$\mathrm{E}\Big[|\lambda_j - \gamma_j^{m,n}|^2\Big] \leqslant C\frac{\log n}{n^2},$$

C depending only on A_1, A_2 and η, which is the claim. The proof of Theorem 5 is complete. □

1.2 Between the bulk and the edge of the spectrum

The aim of this section is to prove an analogous theorem for some eigenvalues between the bulk and the right edge of the spectrum, i.e. for λ_j such that $(1-\eta)n \leqslant j \leqslant n - K\log n$. The precise statement is the following.

Theorem 7. *There exists a constant $\kappa > 0$ (depending on A_1 and A_2) such that the following holds. For all $K > \kappa$, for all $0 < \eta \leqslant \frac{1}{2}$, there exists a constant $C > 0$ (depending on K, η, A_1 and A_2) such that for all covariance matrix $S_{m,n}$, for all $(1-\eta)n \leqslant j \leqslant n - K\log n$,*

$$\mathrm{E}[(\lambda_j - \gamma_j^{m,n})^2] \leqslant C\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}. \tag{18}$$

In particular,

$$\mathrm{Var}(\lambda_j) \leqslant C\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}. \tag{19}$$

As for eigenvalues in the bulk, the proof relies on the determinantal structure of LUE matrices. Recall that this structure together with a bound on the mean counting function (9) leads to the following deviation inequality for the counting function $\mathcal{N}_t$.

$$\mathrm{P}\Big(|\mathcal{N}_t - n\mu_t| \geqslant u + C_1\Big) \leqslant 2\exp\Big(-\frac{u^2}{2\sigma_t^2 + u}\Big).$$

Among Su's results (see [26]), for every $\tilde{\delta} > 0$, for every $\tilde{K} > 0$, there exists $c_{\tilde{\delta},\tilde{K}} > 0$ such that for all t satisfying $0 < b_{m,n} - t < \tilde{\delta}$ and $n(b_{m,n} - t)^{3/2} \geqslant \tilde{K}\log n$,

$$\sigma_t^2 \leqslant c_{\tilde{\delta},\tilde{K}} \log n(b_{m,n} - t)^{3/2}. \tag{20}$$

Combining (10) and (20), deviation inequalities for individual intermediate eigenvalues are then available.

Proposition 8. *Assume that $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2$. There exists $\kappa > 0$ depending only on A_1 and A_2 such that the following holds. Let $K > \kappa$ and $0 < \eta \leqslant \frac{1}{2}$. Then there exist positive constants C, c, C' and c' such that the following holds. For any LUE matrix $S_{m,n}$, for all $(1-\eta)n \leqslant j \leqslant n - K\log n$, for all $c \leqslant u \leqslant c'n$,*

$$\mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| \geqslant \frac{u}{n^{2/3}(n-j)^{1/3}}\Big) \leqslant 4\exp\Big(-\frac{C^2u^2}{C'\log(n-j) + Cu}\Big). \tag{21}$$

The constants C, C' and c' depend only on K, η and A_2, whereas the constant c depends only on K, η, A_1 and A_2.

Note that this proposition still holds when $m = n$, for eigenvalues on the right-side of the spectrum. The proof of Proposition 8 is very similar to what was done for eigenvalues in the bulk. Therefore some details are not reproduced.

Proof. Let $\eta > 0$, $K > 0$ and $u \geqslant 0$. Assume that $(1-\eta)n \leqslant j \leqslant n - K\log n$. Set $u_{n,j} = \frac{u}{n^{2/3}(n-j)^{1/3}}$. As for the bulk case, we start with estimating the probability that λ_j is greater than $\gamma_j^{m,n} + u_{n,j}$. We get

$$\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n} + u_{n,j}\Big) \leqslant \mathrm{P}\Big(|\mathcal{N}_{\gamma_j^{m,n}+u_{n,j}} - n\mu_{\gamma_j^{m,n}+u_{n,j}}| > n\big(\mu_{\gamma_j^{m,n}+u_{n,j}} - \mu_{\gamma_j^{m,n}}\big)\Big).$$

Furthermore,

$$\mu_{\gamma_j^{m,n}+u_{n,j}} - \mu_{\gamma_j^{m,n}} \geqslant \frac{\sqrt{\gamma_j^{m,n} - a_{m,n}}}{3\pi b_{m,n}}(b_{m,n} - \gamma_j^{m,n})^{1/2}u_{n,j},$$

if $\gamma_j^{m,n} + u_{n,j} \leqslant b_{m,n}$. From (13),

$$b_{m,n} - \gamma_j^{m,n} \geqslant \Big(\frac{3\pi}{\sqrt{b_{m,n} - a_{m,n}}}\gamma_j^{m,n}\big(\frac{n-j}{n}\big)\Big)^{2/3}.$$

Moreover, as $\eta \leqslant \frac{1}{2}$, from (14),

$$\sqrt{\gamma_j^{m,n}} \geqslant \sqrt{\gamma_j^{m,n} - a_{m,n}} \geqslant \frac{\pi}{2\sqrt{b_{m,n} - a_{m,n}}}.$$

Therefore

$$b_{m,n} - \gamma_j^{m,n} \geqslant \left(\tfrac{3}{4}\right)^{2/3} \frac{\pi^2}{b_{m,n} - a_{m,n}} \left(\frac{n-j}{n}\right)^{2/3}, \tag{22}$$

and

$$\mu_{\gamma_j^{m,n} + u_{n,j}} - \mu_{\gamma_j^{m,n}} \geqslant \frac{C}{b_{m,n}(b_{m,n} - a_{m,n})} \frac{u}{n},$$

where $C > 0$ is a universal constant. As $\frac{m}{n} \leqslant A_2$,

$$\mu_{\gamma_j^{m,n} + u_{n,j}} - \mu_{\gamma_j^{m,n}} \geqslant \frac{C}{4\sqrt{A_2}(1 + \sqrt{A_2})^2} \frac{u}{n} = C(A_2)\frac{u}{n}.$$

Similarly to the bulk case, we get

$$\mathrm{P}(\lambda_j > \gamma_j^{m,n} + u_{n,j}) \leqslant 2 \exp\Big(- \frac{u^2}{2\sigma^2_{\gamma_j^{m,n} + u_{n,j}} + u} \Big).$$

This relation holds if $c \leqslant u \leqslant n^{2/3}(n-j)^{1/3}(b_{m,n} - \gamma_j^{m,n})$, with c depending only on A_1 and A_2. Let $\alpha \in (0,1)$. Set $c' = \alpha(\frac{3}{4})^{2/3} \frac{\pi^2}{4\sqrt{A_2}}$, depending only on α and A_2. If $u \leqslant c'(n-j)$, then, due to (22), the preceding relation holds. The bound (20) on $\sigma^2_{t_n}$ obtained by Su holds when $0 < b_{m,n} - t_n \leqslant \tilde{\delta}$ and $n(b_{m,n} - t_n)^{3/2} \geqslant \tilde{K} \log n$. Set $t_n = \gamma_j^{m,n} + u_{n,j}$. As $u \geqslant 0$, $0 < b_{m,n} - t_n \leqslant b_{m,n} - \gamma_j^{m,n}$. Therefore, as $j \geqslant (1-\eta)n$, similar computations as for (15) lead to

$$b_{m,n} - t_n \leqslant \Big(\frac{3b_{m,n}\sqrt{b_{m,n} - a_{m,n}}}{1-\eta} \eta \Big)^{2/3} \leqslant \Big(\frac{6(A_2)^{1/4}(1 + \sqrt{A_2})^2}{1-\eta} \eta \Big)^{2/3} = \tilde{\delta}$$

for all n. Moreover,

$$\begin{aligned} n(b_{m,n} - t_n)^{3/2} &= n(b_{m,n} - \gamma_j^{m,n})^{3/2} \Big(1 - \frac{u_{n,j}}{b_{m,n} - \gamma_j^{m,n}} \Big)^{3/2} \\ &\geqslant \frac{3\pi^3}{4(b_{m,n} - a_{m,n})^{3/2}} (n-j) \Big(1 - \frac{c'(n-j)^{2/3}}{n^{2/3}(b_{m,n} - \gamma_j^{m,n})} \Big)^{3/2} \\ &\geqslant \frac{3\pi^3}{4(4\sqrt{A_2})^{3/2}} (1-\alpha)^{3/2} K \log n \\ &\geqslant \tilde{K} \log n, \end{aligned}$$

where $\tilde{K} = \frac{3\pi^3}{4(4\sqrt{A_2})^{3/2}}(1-\alpha)^{3/2}K > 0$. From (20), for all $c \leqslant u \leqslant c'(n-j)$,

$$\mathrm{Var}(\mathcal{N}_{\gamma_j^{m,n}+u_{n,j}}) \leqslant c_{\tilde{\eta},\tilde{K}} \log\Big(n(b_{m,n}-t_n)^{3/2}\Big).$$

But

$$n(b_{m,n}-t_n)^{3/2} \leqslant n\Big(b_{m,n}-\gamma_j^{m,n}\Big)^{3/2}.$$

Using the same techniques as for (13), it is possible to show that

$$(b_{m,n}-\gamma_j^{m,n})^{3/2} \leqslant 6b_{m,n}\sqrt{b_{m,n}-a_{m,n}}\frac{n-j}{n} \leqslant 12(A_2)^{1/4}\Big(1+\sqrt{A_2}\Big)^2\frac{n-j}{n}.$$

Hence $\log\Big(n(b_{m,n}-t_n)^{3/2}\Big) \leqslant \log(n-j)+\log(12(A_2)^{1/4}(1+\sqrt{A_2})^2)$. For $K > \kappa$ with κ large enough depending only on A_2 and for $n \geqslant 2$, $n-j \geqslant K\log n \geqslant 12(A_2)^{1/4}(1+\sqrt{A_2})^2$ and $\mathrm{Var}(\mathcal{N}_{\gamma_j^{m,n}+u_{n,j}}) \leqslant 2c_{\tilde{\delta},\tilde{K}}\log(n-j)$. Therefore

$$\mathrm{P}\Big(\lambda_j > \gamma_j^{m,n}+u_{n,j}\Big) \leqslant 2\exp\Big(-\frac{C^2u^2}{4c_{\tilde{\delta},\tilde{K}}\log(n-j)+Cu}\Big).$$

The proof is concluded similarly to Proposition 6.

□

We turn now to the proof of Theorem 7, in which some details are skipped, due to the similarity with the proof of Theorem 5.

Proof of Theorem 7. Setting $Z = n^{2/3}(n-j)^{1/3}|\lambda_j - \gamma_j^{m,n}|$,

$$\begin{aligned}
\mathrm{E}[Z^2] &= \int_0^\infty \mathrm{P}(Z \geqslant v)2v\,dv \\
&= \int_0^c \mathrm{P}(Z \geqslant v)2v\,dv + \int_c^{\frac{C'}{C}\log(n-j)} \mathrm{P}(Z \geqslant v)2v\,dv \\
&\quad + \int_{\frac{C'}{C}\log(n-j)}^{c'(n-j)} \mathrm{P}(Z \geqslant v)2v\,dv + \int_{c'(n-j)}^\infty \mathrm{P}(Z \geqslant v)2v\,dv \\
&\leqslant c^2 + J_1 + J_2 + J_3,
\end{aligned}$$

where c, c', C and C' are given by Proposition 8. Repeating the computations carried out with I_2 in the proof of Theorem 5 yields

$$\begin{aligned}
J_3 &\leqslant 2n^{4/3}(n-j)^{2/3}\sqrt{\mathrm{E}[(\lambda_j-\gamma_j^{m,n})^4]}\exp\Big(-\frac{1}{2}\frac{C^2C'^2(n-j)^2}{C'\log(n-j)+Cc'(n-j)}\Big) \\
&\leqslant 2An^4\exp\Big(-\frac{1}{2}\frac{C^2c'^2(n-j)^2}{C'\log(n-j)+Cc'(n-j)}\Big),
\end{aligned}$$

where $A > 0$ is a numerical constant. The last inequality is due to (17). For n large enough (depending on η, A_2 and K), $C' \log(n-j) \leqslant Cc'(n-j)$ and

$$J_3 \leqslant 2A \exp\Big(4\log n - \frac{Cc'}{4}(n-j)\Big).$$

Then, as $n - j \geqslant K \log n$,

$$J_3 \leqslant 2A \exp\Big(\Big(4 - \frac{KCc'}{4}\Big)\log n\Big).$$

Recall from the proof of Proposition 8 that $c' = \alpha c'(A_2)$ where $\alpha \in (0,1)$ is a universal constant and $c'(A_2)$ depends only on A_2. Furthermore, the constant C depends only on A_2. Therefore, if we choose $\kappa > 0$ such that $\kappa > \frac{16}{Cc'(A_2)}$, then $\frac{KCc'}{4} > 4$. The right-hand side goes thus to 0 when n goes to infinity. As a consequence, there exists a constant $C > 0$ depending only on A_2, η and K such that

$$J_3 \leqslant C.$$

The integral J_1 is handled as I_1, using that, in the range $v \leqslant \frac{C'}{C}\log(n-j)$,

$$P(Z \geqslant v) \leqslant 4\exp\Big(-\frac{B}{\log(n-j)}v^2\Big),$$

where B depends only on K, η and A_2 (this is due to Proposition 8). Hence, there exists a constant C depending only on A_2, η and K such that

$$J_1 \leqslant C\log(n-j).$$

Finally, J_2 is handled similarly. In the range $\frac{C'}{C}\log(n-j) \leqslant v \leqslant c'(n-j)$, from Proposition 8,

$$P(Z \geqslant v) \leqslant 4\exp\Big(-\frac{C}{2}v\Big).$$

Thus

$$J_2 \leqslant 4\int_{\frac{C'}{C}\log(n-j)}^{c'(n-j)} \exp\Big(-\frac{C}{2}v\Big)2v\,dv \leqslant 4\int_0^{\infty} \exp\Big(-\frac{C}{2}v\Big)2v\,dv.$$

Then J_2 is bounded by a constant, which depends only on A_2. There exists thus a constant $C > 0$ such that

$$J_2 \leqslant C.$$

Summarizing the previous steps, $\mathrm{E}[Z^2] \leqslant C\log(n-j)$, where C depends only on A_1, A_2, η and K. Therefore

$$\mathrm{E}\Big[|\lambda_j - \gamma_j|^2\Big] \leqslant C\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}},$$

which is the claim. $\square$

1.3 At the edge of the spectrum

In [18], Ledoux and Rider gave unified proofs of precise small deviation inequalities for the largest eigenvalues of β-ensembles. The results hold in particular for LUE matrices ($\beta = 2$) and for LOE matrices ($\beta = 1$). The following theorem summarizes some of the relevant inequalities for the LUE.

Theorem 9. *[18] Let $A_1 > 1$. There exists a constant $C > 0$ depending only on A_1 such that the following holds. Let $S_{m,n}$ be a LUE matrix. Denote by λ_n the maximal eigenvalue of $S_{m,n}$. Then, for all $n \in \mathbb{N}$, for all $m \in \mathbb{N}$ such that $m > A_1 n$ and for all $0 < \varepsilon \leqslant 1$,*

$$\mathrm{P}\Big(\lambda_n \leqslant b_{m,n}(1-\varepsilon)\Big) \leqslant C^2 \exp\Big(-\frac{2}{C} n^2 \varepsilon^3\Big), \tag{23}$$

and

$$\mathrm{P}\Big(\lambda_n \geqslant b_{m,n}(1+\varepsilon)\Big) \leqslant C \exp\Big(-\frac{2}{C} n \varepsilon^{3/2}\Big). \tag{24}$$

The large deviation tails are also known. The following corollary can be deduced by integrating these inequalities.

Corollaire 10. *[18] Let $S_{m,n}$ be a LUE matrix. Then there exists a universal constant $C > 0$ such that for all $n \geqslant 1$, for all $m \in \mathbb{N}$ such that $m > A_1 n$,*

$$\mathrm{Var}(\lambda_n) \leqslant \mathrm{E}\Big[(\lambda_n - b_{m,n})^2\Big] \leqslant C n^{-4/3}.$$

Similar results are probably true for the k^{th} largest eigenvalue (for $k \in \mathbb{N}$ fixed). The authors established also a left-side deviation inequality for the smallest eigenvalue in the case when $m > A_1 n$.

$$\mathrm{P}\Big(\lambda_1 \leqslant a_{m,n}(1-\varepsilon)\Big) \leqslant C \exp\Big(-\frac{2}{C} n \varepsilon^{3/2}\Big), \tag{25}$$

for all $0 < \varepsilon \leqslant 1$. But no right-side deviation inequality seems to be known for the smallest eigenvalue λ_1 and therefore we cannot deduce a precise bound on the variance of the smallest eigenvalue.

2 Variance bounds for families of covariance matrices

The previously achieved bounds on the variance of eigenvalues for LUE matrices are then extended to families of more general covariance matrices. It is due to the combination of two very recent results, some localization properties established by Pillai and Yin [23] and to the Four Moment Theorem proved by Tao and Vu [28] and Wang [33].

2.1 Localization properties and the Four Moment Theorem

This subsection is devoted to the statement of the previously mentioned results which will be used in order to extend variance bounds to large families of non Gaussian covariance matrices. Matrices which are considered in this section are covariance matrices $S_{m,n}$ satisfying condition $(C0)$, defined by the following. Say that $S_{m,n}$ satisfies condition $(C0)$ if its entries X_{ij} are independent and have an exponential decay: there are positive constants B_1 and B_2 such that

$$\forall\, i \in \{1, \dots, n\}, \forall\, j \in \{1, \dots, m\},\ \mathrm{P}\Big(|X_{ij}| \geqslant t^{B_1}\Big) \leqslant e^{-t}$$

for all $t \geqslant B_2$.

Pillai and Yin proved in [23] a Localization Theorem similar to the one proved by Erdös, Yau and Yin in [9]. This theorem establishes that the eigenvalues are highly localized around their theoretical locations $\gamma_j^{m,n}$.

Theorem 11 (Localization [23])**.** *Let $S_{m,n}$ be a random covariance matrix whose entries satisfy condition $(C0)$. Suppose that $1 < A_1 \leqslant \frac{m}{n} \leqslant A_2 < +\infty$. There are positive universal constants c and C such that, for any $1 \leqslant j \leqslant n$,*

$$\mathrm{P}\left(|\lambda_j - \gamma_j^{m,n}| \geqslant (\log n)^{C \log\log n} n^{-2/3} \min(j, n+1-j)^{-1/3}\right) \leqslant Ce^{-(\log n)^{c \log\log n}}. \quad (26)$$

This deviation inequality (26) can be used to reach an almost optimal bound on the variance. Indeed, due to (26) and the Cauchy-Schwarz inequality, $\mathrm{Var}(\lambda_j)$ may be bounded by $\frac{(\log n)^{2C\log\log n}}{n^2}$ in the bulk of the spectrum, which is almost the right order for the variance. In order to remove the $\log\log n$ term, we turn now to the Four Moment Theorem. This theorem was proved for the bulk of the spectrum by Tao and Vu [28] and extended to the edge by Wang [33]. From now, we consider covariance matrices $S_{m,n}$ which satisfy condition $(C0)$ and whose entries match the entries of a LUE matrix up to order 4. Say that two complex random variables ξ and ξ' match to order k if

$$\mathrm{E}\Big[\Re(\xi)^m \Im(\xi)^l\Big] = \mathrm{E}\Big[\Re(\xi')^m \Im(\xi')^l\Big]$$

for all $m, l \geqslant 0$ such that $m + l \leqslant k$.

Theorem 12 (Four Moment Theorem [28, 33])**.** *There exists a small positive constant c_0 such that the following holds. Let $S_{m,n} = \frac{1}{n}X^*X$ and $S'_{m,n} = \frac{1}{n}X'^*X'$ be two random covariance matrices satisfying condition $(C0)$. Assume that, for $1 \leqslant i \leqslant n$, X_{ij} and X'_{ij} match to order 4. Let $G : \mathbb{R} \to \mathbb{R}$ be a smooth function such that:*

$$\forall\ 0 \leqslant k \leqslant 5, \quad \forall x \in \mathbb{R}, \quad \Big|G^{(k)}(x)\Big| \leqslant n^{c_0}. \quad (27)$$

Then, for all $1 \leqslant i \leqslant n$ and for n large enough (depending on constants B_1 and B_2 in condition $(C0)$),

$$\left|\mathrm{E}[G(n\lambda_i)] - \mathrm{E}[G(n\lambda_i')]\right| \leqslant n^{-c_0}. \tag{28}$$

Suppose Theorem 12 apply with $G_j : x \in \mathbb{R} \mapsto (x - n\gamma_j^{m,n})^2$. Then (12) writes

$$\left|n^2\mathrm{E}[(\lambda_j - \gamma_j^{m,n})^2] - n^2\mathrm{E}[(\lambda_j' - \gamma_j^{m,n})^2]\right| \leqslant n^{-c_0}.$$

As $\mathrm{E}[(\lambda_j' - \gamma_j^{m,n})^2]$ is bounded by $\frac{\log n}{n^2}$, $\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}$ or $n^{-4/3}$, which are bigger than n^{-2-c_0}, the bounds could be extended. Unfortunately, G_j does not satisfy (27). To get round this difficulty, the Four Moment Theorem 12 is applied to a smooth truncation of G_j. The Localization Theorem 11 provides a small area around $\gamma_j^{m,n}$ where λ_j is very likely to be in so that the error due to the truncation is well controlled. Details are contained in the following subsection.

2.2 Comparison with LUE matrices

Let $S_{m,n}$ be a covariance matrix and $S'_{m,n}$ be a LUE matrix such that they satisfy the hypotheses of Theorem 12. Note that the following procedure is valid for eigenvalues in the bulk and at the edge of the spectrum, as well as for intermediate eigenvalues.

Let $1 \leqslant j \leqslant n$. Set $R_n^{(j)} = (\log n)^{C \log\log n} n^{1/3} \min(j, n+1-j)^{-1/3}$ and $\varepsilon_n = Ce^{-(\log n)^{c \log\log n}}$. Then Theorem 11 leads to:

$$\mathrm{P}\left(|\lambda_j - \gamma_j^{m,n}| \geqslant \frac{R_n^{(j)}}{n}\right) \leqslant \varepsilon_n. \tag{29}$$

Let ψ be a smooth function with support $[-2, 2]$ and values in $[0, 1]$ such that $\psi(x) = \frac{1}{10}x^2$ for all $x \in [-1; 1]$. Set $G_j : x \in \mathbb{R} \mapsto \psi\Big(\frac{x - n\gamma_j}{R_n^{(j)}}\Big)$. We want to apply Tao and Vu's Four Moment Theorem 12 to G_j. As ψ is smooth and has compact support, its first five derivatives are bounded by $M > 0$. Then, for all $0 \leqslant k \leqslant 5$, for all $x \in \mathbb{R}$,

$$\left|G_j^{(k)}(x)\right| \leqslant \frac{M}{(R_n^{(j)})^k} \leqslant n^{c_0},$$

where the last inequality holds for n large enough (depending only on M and c_0). Then, the Four Moment Theorem 12 yields:

$$\left|\mathrm{E}[G_j(n\lambda_j)] - \mathrm{E}[G_j(n\lambda_j')]\right| \leqslant n^{-c_0} \tag{30}$$

for large enough n. But

$$\begin{aligned}
&\mathrm{E}[G_j(n\lambda_j)] \\
&\quad = \tfrac{1}{10}\mathrm{E}\Big[\Big(\tfrac{n\lambda_j - n\gamma_j^{m,n}}{R_n^{(j)}}\Big)^2 1_{\frac{|n\lambda_j - n\gamma_j^{m,n}|}{R_n^{(j)}}\leqslant 1}\Big] + \mathrm{E}\Big[G_j(n\lambda_j)1_{\frac{|n\lambda_j - n\gamma_j^{m,n}|}{R_n^{(j)}}>1}\Big] \\
&\quad = \frac{n^2}{10(R_n^{(j)})^2}\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2 1_{|\lambda_j - \gamma_j^{m,n}|\leqslant \frac{R_n^{(j)}}{n}}\Big] + \mathrm{E}\Big[G_j(n\lambda_j)1_{\frac{|n\lambda_j - n\gamma_j^{m,n}|}{R_n^{(j)}}>1}\Big].
\end{aligned}$$

On the one hand,

$$\begin{aligned}
\mathrm{E}\Big[G_j(n\lambda_j)1_{\frac{|n\lambda_j - n\gamma_j^{m,n}|}{R_n^{(j)}}>1}\Big] &\leqslant \mathrm{P}\Big(|n\lambda_j - n\gamma_j^{m,n}| > R_n^{(j)}\Big) \\
&\leqslant \mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| > \frac{R_n^{(j)}}{n}\Big) \\
&\leqslant \varepsilon_n.
\end{aligned}$$

On the other hand,

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2 1_{|\lambda_j - \gamma_j^{m,n}|\leqslant \frac{R_n^{(j)}}{n}}\Big] = \mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] - \mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2 1_{|\lambda_j - \gamma_j^{m,n}|> \frac{R_n^{(j)}}{n}}\Big].$$

As condition $(C0)$ is satisfied, the 8-th moment of the entries is uniformly bounded by a constant which depends only on constants B_1 and B_2 in condition $(C0)$. Then, from the Cauchy-Schwarz inequality and (16),

$$\begin{aligned}
\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2 1_{|\lambda_j - \gamma_j^{m,n}|> \frac{R_n^{(j)}}{n}}\Big] &\leqslant \sqrt{\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^4\Big]\mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| > \tfrac{R_n^{(j)}}{n}\Big)} \\
&\leqslant An^2\sqrt{\varepsilon_n}
\end{aligned}$$

where $A > 0$ is a numerical constant. Then

$$\begin{aligned}
\mathrm{E}\Big[G_j(n\lambda_j)\Big] &= \frac{n^2}{10\Big(R_n^{(j)}\Big)^2}\Big(\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] + O\Big(n^2\varepsilon_n^{1/2}\Big)\Big) + O(\varepsilon_n) \\
&= \frac{n^2}{10\Big(R_n^{(j)}\Big)^2}\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] + O\Big(n^4\varepsilon_n^{1/2}\Big(R_n^{(j)}\Big)^{-2}\Big) + O(\varepsilon_n).
\end{aligned}$$

Repeating the same computations gives similarly

$$\mathrm{E}\Big[G_j(n\lambda'_j)\Big] = \frac{n^2}{10\Big(R_n^{(j)}\Big)^2}\mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] + O\Big(n^4\varepsilon_n^{1/2}\Big(R_n^{(j)}\Big)^{-2}\Big) + O(\varepsilon_n).$$

Then (30) yields

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] = \mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] + O\Big(n^2\varepsilon_n^{1/2} + n^{-2}(R_n^{(j)})^2\varepsilon_n + (R_n^{(j)})^2 n^{-c_0-2}\Big).$$

As the first two error terms are smaller than the third one, the preceding equation becomes

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] = \mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] + O\Big((R_n^{(j)})^2 n^{-c_0-2}\Big). \tag{31}$$

2.3 Variance bounds

(31) is true for all eigenvalue λ_j. We estimate the error term $O\Big((R_n^{(j)})^2 n^{-c_0-2}\Big)$ differently according to the location of the eigenvalue in the spectrum, in order to get the announced bounds.

2.3.1 Inside the bulk of the spectrum

Let $0 < \eta \leqslant \frac{1}{2}$ and $\eta n \leqslant j \leqslant (1-\eta)n$. From Theorem 5, $\mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] \leqslant C\frac{\log n}{n^2}$. Thus, from (31), it remains to show that the error term is smaller than $\frac{\log n}{n^2}$. But

$$R_n^{(j)} = (\log n)^{C\log\log n} n^{1/3}\min(j, n+1-j)^{-1/3} \leqslant \eta^{-1/3}(\log n)^{C\log\log n}.$$

Then $(R_n^{(j)})^2 n^{-c_0-2} = o_\eta\Big(\frac{\log n}{n^2}\Big)$. As a consequence,

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] = \mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] + o_\eta\Big(\frac{\log n}{n^2}\Big)$$

and we get the desired result

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] \leqslant C\frac{\log n}{n^2},$$

C depending only on η, A_1 and A_2.

2.3.2 Between the bulk and the edge of the spectrum

Let $0 < \eta \leqslant \frac{1}{2}$, $K > \kappa$ and $(1-\eta)n \leqslant j \leqslant n - K\log n$. From Theorem 7, $\mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] \leqslant C\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}$. Thus, from (31), it remains to show that the error term is smaller than $\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}$. But

$$R_n^{(j)} = (\log n)^{C\log\log n} n^{1/3}(n+1-j)^{-1/3}.$$

Then $(R_n^{(j)})^2 n^{-c_0-2} = o\Big(\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}\Big)$. As a consequence,

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] = \mathrm{E}\Big[(\lambda'_j - \gamma_j^{m,n})^2\Big] + o\Big(\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}}\Big)$$

and we get the desired result

$$\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] \leqslant C\frac{\log(n-j)}{n^{4/3}(n-j)^{2/3}},$$

C depending only on η, A_1, A_2 and K. A similar result probably holds for the left-side of the spectrum, when $\rho > 1$.

2.3.3 At the edge of the spectrum

From Corollary 10, $\mathrm{E}\Big[(\lambda'_n - \gamma_n^{m,n})^2\Big] = \mathrm{E}\Big[(\lambda'_n - b_{m,n})^2\Big] \leqslant Cn^{-4/3}$. By means of (31), it remains to prove that the error term is smaller than $n^{-4/3}$. We have

$$R_n^{(n)} = (\log n)^{C \log\log n} n^{1/3}.$$

Consequently $(R_n^{(n)})^2 n^{-c_0-2} = o\Big(n^{-4/3}\Big)$. Then

$$\mathrm{E}\Big[(\lambda_n - b_{m,n})^2\Big] = \mathrm{E}\Big[(\lambda'_n - 2)^2\Big] + o\Big(n^{-4/3}\Big)$$

and

$$\mathrm{E}\Big[(\lambda_n - 2)^2\Big] \leqslant Cn^{-4/3}.$$

If this bound holds for the smallest eigenvalue λ_1 of LUE matrices, the same result is available for non Gaussian covariance matrices.

3 Real Wishart matrices

The aim of this section is to prove Theorems 1, 2 and 3 for real covariance matrices. The Four Moment Theorem (Theorem 12) by Tao, Vu and Wang as well as Pillai and Yin's Localization Theorem (Theorem 11) still hold for real covariance matrices. Section 2 is therefore valid for real matrices. The point is then to establish the results in the LOE case.

As announced in Section 1.3, the variance of eigenvalues at the right edge of the spectrum is known to be bounded by $n^{-4/3}$ for LOE matrices (see [18]). The conclusion for the largest eigenvalue is then established for large families of real covariance matrices.

$$\mathrm{Var}(\lambda_n) \leqslant \frac{\tilde{C}}{n^{4/3}}.$$

For eigenvalues in the bulk of the spectrum, following O'Rourke's approach (see [20]), a Central Limit Theorem similar to the one established by Su in [26] may be proved. In particular, the normalization is still of the order of $(\frac{\log n}{n^2})^{1/2}$ and differs from the complex case only by a constant. It is therefore natural to expect the same bound on the variance for LOE matrices. The situation is completely similar for intermediate eigenvalues. But LOE matrices do not have the same determinantal properties as LUE matrices, and it is therefore not clear that a deviation inequality (similar to (10)) holds for the eigenvalue counting function. However, LOE and LUE matrices are linked by interlacing formulas established by Forrester and Rains (see [11]). These formulas lead to the following relation between the eigenvalue counting functions in the complex and real cases: for all $t \in \mathbb{R}$,

$$\mathcal{N}_t(S^{\mathbb{C}}_{m,n}) \overset{(d)}{=} \frac{1}{2}\Big(\mathcal{N}_t(S^{\mathbb{R}}_{m,n}) + \mathcal{N}_t(\tilde{S}^{\mathbb{R}}_{m,n})\Big) + \zeta_N(t), \tag{32}$$

where $S^{\mathbb{C}}_{m,n}$ is from the LUE, $S^{\mathbb{R}}_{m,n}$, $\tilde{S}^{\mathbb{R}}_{m,n}$ are independent matrices from the LOE and $\zeta_N(t)$ takes values in $\big\{-\frac{3}{2}, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, \frac{3}{2}\big\}$. See [20] for more details.

The aim is now to establish a deviation inequality for the eigenvalue counting function similar to (10). From (10), we know that for all $u \geqslant 0$,

$$\mathrm{P}\Big(|\mathcal{N}_t(S^{\mathbb{C}}_{m,n}) - n\mu_t| \geqslant u + C_1\Big) \leqslant 2\exp\Big(-\frac{u^2}{2\sigma_t^2 + u}\Big).$$

Set $C_1' = C_1 + \frac{3}{2}$ and let $u \geqslant 0$. We can then write

$$\begin{aligned}
\mathrm{P}&\Big(\mathcal{N}_t(S^{\mathbb{R}}_{m,n}) - n\mu_t \geqslant u + C_1'\Big)^2 \\
&= \mathrm{P}\Big(\mathcal{N}_t(S^{\mathbb{R}}_{m,n}) - n\mu_t \geqslant u + C_1',\ \mathcal{N}_t(\tilde{S}^{\mathbb{R}}_{m,n}) - n\mu_t \geqslant u + C_1'\Big) \\
&\leqslant \mathrm{P}\Big(\tfrac{1}{2}\Big(\mathcal{N}_t(S^{\mathbb{R}}_{m,n}) + \mathcal{N}_t(\tilde{S}^{\mathbb{R}}_{m,n})\Big) - n\mu_t \geqslant u + C_1'\Big) \\
&\leqslant \mathrm{P}\Big(\mathcal{N}_t(S^{\mathbb{C}}_{m,n}) - n\mu_t \geqslant u + C_1' - \frac{3}{2}\Big) \\
&\leqslant 2\exp\Big(-\frac{u^2}{2\sigma_t^2 + u}\Big).
\end{aligned}$$

Repeating the computations for $\mathrm{P}\Big(\mathcal{N}_t(S^{\mathbb{R}}_{m,n}) - n\mu_t \leqslant -u - C_1'\Big)$ and combining with the preceding yield

$$\mathrm{P}\Big(|\mathcal{N}_t(S^{\mathbb{R}}_{m,n}) - n\mu_t| \geqslant u + C_1'\Big) \leqslant 2\sqrt{2}\exp\Big(-\frac{u^2}{4\sigma_t^2 + 2u}\Big). \tag{33}$$

Note that σ_t^2 is still the variance of $\mathcal{N}_t(S^{\mathbb{C}}_{m,n})$ in the preceding formula.

What remains then to be proved is very similar to the complex case. From (33) and Su's bounds on the variance σ_t^2 (see (11) for the bulk case and (20) for the intermediate case), deviation inequalities for individual eigenvalues can be deduced, as was done to prove Propositions 6 and 8. It is then straightforward to derive the announced bounds on the variances for LOE matrices. The argument developed in Section 2 in order to extend the LUE results to large families of covariance matrices can be reproduced to reach the desired bounds on the variances of eigenvalues in the bulk and between the bulk and the edge of the spectrum for families of real covariance matrices.

4 Rate of convergence towards the Marchenko-Pastur distribution

In this whole section, we suppose that Theorem 2 holds for left-side intermediate eigenvalues. The bounds on $\mathbb{E}[(\lambda_j - \gamma_j^{m,n})^2]$ developed in the preceding sections lead to a bound on the rate of convergence of the empirical spectral measure $L_{m,n}$ towards the Marchenko-Pastur distribution in terms of 2-Wasserstein distance. Recall that $W_2(L_{m,n}, \mu_{m,n})$ is a random variable defined by

$$W_2(L_{m,n}, \mu_{m,n}) = \inf \left(\int_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2},$$

where the infimum is taken over all probability measures π on $\mathbb{R}^2$ with respective marginals $L_{m,n}$ and $\mu_{m,n}$. To achieve the expected bound, we rely on another expression of W_2 in terms of distribution functions, namely

$$W_2^2(L_{m,n}, \mu_{m,n}) = \int_0^1 \left(F_{m,n}^{-1}(x) - G_{m,n}^{-1}(x) \right)^2 dx, \tag{34}$$

where $F_{m,n}^{-1}$ (respectively $G_{m,n}^{-1}$) is the generalized inverse of the distribution function $F_{m,n}$ (respectively $G_{m,n}$) of $L_{m,n}$ (respectively $\mu_{m,n}$) (see for example [32]). On the basis of this representation, the following statement may be derived.

Proposition 13. *There exists a constant $C > 0$ depending only on A_2 such that for all $1 \leqslant \frac{m}{n} \leqslant A_2$,*

$$W_2^2(L_{m,n}, \mu_{m,n}) \leqslant \frac{2}{n} \sum_{j=1}^{n} (\lambda_j - \gamma_j^{m,n})^2 + \frac{C}{n^2}. \tag{35}$$

Proof. From (34),

$$W_2^2(L_{m,n}, \mu_{m,n}) = \int_0^1 \left(F_{m,n}^{-1}(x) - G_{m,n}^{-1}(x) \right)^2 dx.$$

Then,

$$\begin{aligned}W_2^2(L_{m,n},\mu_{m,n}) &= \sum_{j=1}^{n}\int_{\frac{j-1}{n}}^{\frac{j}{n}}\Big(\lambda_j - G_{m,n}^{-1}(x)\Big)^2\,dx\\ &\leqslant \frac{2}{n}\sum_{j=1}^{n}(\lambda_j-\gamma_j^{m,n})^2 + 2\sum_{j=1}^{n}\int_{\frac{j-1}{n}}^{\frac{j}{n}}\Big(\gamma_j^{m,n} - G_{m,n}^{-1}(x)\Big)^2\,dx.\end{aligned}$$

But $\gamma_j^{m,n} = G_{m,n}^{-1}\big(\frac{j}{n}\big)$ and $G_{m,n}^{-1}$ is non-decreasing. Therefore, $\Big|\gamma_j^{m,n} - G_{m,n}^{-1}(x)\Big| \leqslant \gamma_j^{m,n} - \gamma_{j-1}^{m,n}$ for all $x \in \Big[\frac{j-1}{n}, \frac{j}{n}\Big]$. Consequently,

$$W_2^2(L_{m,n},\mu_{m,n}) \leqslant \frac{2}{n}\sum_{j=1}^{n}(\lambda_j-\gamma_j^{m,n})^2 + \frac{2}{n}\sum_{j=1}^{n}(\gamma_j^{m,n}-\gamma_{j-1}^{m,n})^2. \tag{36}$$

But if $j-1 \geqslant \frac{n}{2}$,

$$\begin{aligned}\frac{1}{n} &= \int_{\gamma_{j-1}^{m,n}}^{\gamma_j^{m,n}} \frac{1}{2\pi x}\sqrt{(b_{m,n}-x)(x-a_{m,n})}\,dx\\ &\geqslant \frac{\sqrt{\gamma_{j-1}^{m,n}-a_{m,n}}}{2\pi b_{m,n}}\int_{\gamma_{j-1}^{m,n}}^{\gamma_j^{m,n}}\sqrt{b_{m,n}-x}\,dx\\ &\geqslant \frac{\sqrt{\gamma_{j-1}^{m,n}-a_{m,n}}}{3\pi b_{m,n}}(b_{m,n}-\gamma_{j-1}^{m,n})^{3/2}\left(1-\left(1-\frac{\gamma_j^{m,n}-\gamma_{j-1}^{m,n}}{b_{m,n}-\gamma_{j-1}^{m,n}}\right)^{3/2}\right)\\ &\geqslant \frac{1}{3b_{m,n}\sqrt{b_{m,n}-a_{m,n}}}(b_{m,n}-\gamma_{j-1}^{m,n})^{1/2}(\gamma_j^{m,n}-\gamma_{j-1}^{m,n})\\ &\geqslant C(A_2)\Big(\frac{n-j+1}{n}\Big)^{1/3}(\gamma_j^{m,n}-\gamma_{j-1}^{m,n}),\end{aligned}$$

from (13). Then

$$\gamma_j^{m,n}-\gamma_{j-1}^{m,n} \leqslant \frac{C(A_2)}{n^{2/3}(n-j+1)^{2/3}}.$$

It may be shown that a similar bound holds if $j-1 \leqslant \frac{n}{2}$. As a summary, there exists a constant $c > 0$ depending only on A_2 such that, for all $j \geqslant 2$,

$$\gamma_j^{m,n}-\gamma_{j-1}^{m,n} \leqslant \frac{c}{n^{2/3}\min(j,n+1-j)^{1/3}}. \tag{37}$$

This yields

$$\sum_{j=1}^{n}(\gamma_j^{m,n}-\gamma_{j-1}^{m,n})^2 \leqslant \frac{c^2}{n^{4/3}}\sum_{j=1}^{n}\frac{1}{\min(j,n+1-j)^{2/3}} \leqslant \frac{C}{n},$$

where $C > 0$ depends only on A_2. Then (36) becomes

$$W_2^2(L_{m,n},\mu_{m,n}) \leqslant \frac{2}{n}\sum_{j=1}^{n}(\lambda_j-\gamma_j^{m,n})^2+\frac{C}{n^2},$$

where $C > 0$ depends only on A_2, which is the claim. □

Proof of Corollary 4. Let $n \geqslant 2$. Due to Proposition 13,

$$\mathrm{E}\Big[W_2^2(L_{m,n},\mu_{m,n})\Big] \leqslant \frac{2}{n}\sum_{j=1}^{n}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]+\frac{C}{n^2}.$$

We then make use of the bounds on $\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]$ produced in the previous sections. Set $0<\eta\leqslant\frac{1}{2}$ and $K>\kappa$ so that $K\log n\leqslant \eta n$. We first decompose

$$\begin{aligned}
\sum_{j=1}^{n}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big] &= \sum_{j=1}^{K\log n}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]+\sum_{j=K\log n+1}^{\eta n}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]\\
&+\sum_{j=\eta n+1}^{(1-\eta)n-1}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]+\sum_{j=(1-\eta)n}^{n-K\log n-1}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]\\
&+\sum_{j=n-K\log n}^{n}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]\\
&=\Sigma_1+\Sigma_2+\Sigma_3+\Sigma_4+\Sigma_5.
\end{aligned}$$

The sum Σ_3 will be bounded using the bulk case (Theorem 1), while Theorem 2 will be used to handle Σ_2 and Σ_4. A crude version of Theorem 11 will be enough to bound Σ_1 and Σ_5. To start with thus, from Theorem 1,

$$\Sigma_3 \leqslant \sum_{j=\eta n+1}^{(1-\eta)n-1} C\frac{\log n}{n^2}\leqslant C\frac{\log n}{n}.$$

Secondly, from Theorem 2,

$$\Sigma_2+\Sigma_4 \;\leqslant\; \frac{C}{n^{4/3}}\sum_{j=K\log n+1}^{\eta n}\frac{\log j}{j^{2/3}} \;\leqslant\; C\frac{\log n}{n}.$$

Next Σ_1 and Σ_5 have only $K\log n$ terms. If each term is bounded by $\frac{C}{n}$ where C is a positive universal constant, we get that $\Sigma_1+\Sigma_5\leqslant\frac{2KC\log n}{n}$, which is enough to prove the desired result on $\sum_{j=1}^{n}\mathrm{E}\Big[(\lambda_j-\gamma_j^{m,n})^2\Big]$. For n large enough depending only on constant C in Theorem 11, $\frac{1}{\sqrt{n}}\geqslant\frac{(\log n)^{C\log\log n}}{n^{2/3}\min(j,n+1-j)^{1/3}}$ and Theorem 11 yields

$$\mathrm{P}\bigg(|\lambda_j-\gamma_j^{m,n}|\geqslant\frac{1}{\sqrt{n}}\bigg)\leqslant Ce^{-(\log n)^{c\log\log n}}.$$

Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] &\leqslant \mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2 1_{|\lambda_j - \gamma_j^{m,n}| \leqslant \frac{1}{\sqrt{n}}}\Big] + \mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2 1_{|\lambda_j - \gamma_j^{m,n}| > \frac{1}{\sqrt{n}}}\Big] \\
&\leqslant \frac{1}{n} + \sqrt{\mathrm{E}\Big[|\lambda_j - \gamma_j^{m,n}|^4\Big]}\sqrt{\mathrm{P}\Big(|\lambda_j - \gamma_j^{m,n}| > \frac{1}{\sqrt{n}}\Big)} \\
&\leqslant \frac{1}{n} + \sqrt{3}Cn^2 e^{-(\log n)^{c \log\log n}}.
\end{aligned}$$

As $\sqrt{3}Cn^2 e^{-(\log n)^{c\log\log n}} = o(\frac{1}{n})$, there exists a constant $C > 0$ such that $\mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] \leqslant \frac{C}{n}$. Then

$$\Sigma_1 + \Sigma_5 \leqslant 2KC\frac{\log n}{n}.$$

As a consequence,

$$\sum_{j=1}^{n} \mathrm{E}\Big[(\lambda_j - \gamma_j^{m,n})^2\Big] \leqslant C\frac{\log n}{n}.$$

Therefore

$$\mathrm{E}\Big[W_2(L_{m,n}, \mu_{m,n})^2\Big] \leqslant C\frac{\log n}{n^2},$$

where $C > 0$ depends only on A_1 and A_2, which is the claim. The corollary is thus established. □

Acknowledgements

I would like to thank my advisor, Michel Ledoux, for bringing up this problem to me, for the several discussions we had about this work, and Emmanuel Boissard for very useful conversations about Wasserstein distances.

References

[1] G. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[2] Z. Bai, P. Krishnaiah, and Y. Yin. On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix. *Probab. Theory Related Fields*, 78(4):509–521, 1988.

[3] Z. Bai and J. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

[4] Z. Bai, J. Silverstein, and Y. Yin. A note on the largest eigenvalue of a large-dimensional sample covariance matrix. *J. Multivariate Anal.*, 26(2):166–168, 1988.

[5] Z. Bai and Y. Yin. Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993.

[6] A. Borodin and P. Forrester. Increasing subsequences and the hard-to-soft edge transition in matrix ensembles. *J. Phys. A*, 36(12):2963–2981, 2003. Random matrix theory.

[7] S. Dallaporta. Eigenvalue variance bounds for Wigner and covariance random matrices. *RMTA*, 1(3), 2012.

[8] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[9] L. Erdös, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized wigner matrices. *arXiv:1007.4652*, 2010.

[10] O. Feldheim and S. Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.*, 20(1):88–123, 2010.

[11] P. Forrester and E. Rains. Interrelationships between orthogonal, unitary and symplectic matrix ensembles. In *Random matrix models and their applications*, volume 40 of *Math. Sci. Res. Inst. Publ.*, pages 171–207. Cambridge Univ. Press, Cambridge, 2001.

[12] F. Götze and A. Tikhomirov. Rate of convergence in probability to the Marchenko-Pastur law. *Bernoulli*, 10(3):503–548, 2004.

[13] F. Götze and A. Tikhomirov. On the rate of convergence to the Marchenko-Pastur distribution. *arXiv:1110.1284*, 2011.

[14] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136, 2000.

[15] J. Ben Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. *Probab. Surv.*, 3:206–229, 2006.

[16] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(2):437–476, 2000.

[17] I. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.

[18] M. Ledoux and B. Rider. Small deviations for beta ensembles. *Electron. J. Probab.*, 15(41):1319–1343, 2010.

[19] M. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[20] S. O'Rourke. Gaussian fluctuations of eigenvalues in Wigner random matrices. *J. Stat. Phys.*, 138(6):1045–1066, 2010.

[21] L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*, volume 171 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2011.

[22] S. Péché. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields*, 143(3-4):481–516, 2009.

[23] N. Pillai and J. Yin. Universality of covariance matrices. *arXiv:1110.2501*, 2011.

[24] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians. Volume III*, pages 1576–1602, New Delhi, 2010. Hindustan Book Agency.

[25] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002.

[26] Z. Su. Gaussian fluctuations in complex sample covariance matrices. *Electron. J. Probab.*, 11(48):1284–1320, 2006.

[27] T. Tao and V. Vu. Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20(1):206–297, 2010.

[28] T. Tao and V. Vu. Random covariance matrices: universality of local statistics of eigenvalues. *Annals of Probability*, 40(3):1285–1315, 2012.

[29] A. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1998.

[30] R. Vershynin. How close is the sample covariance matrix to the actual covariance matrix? *J. Theoret. Probab.*, 25(3):655–686, 2012.

[31] R. Vershynin. *Introduction to the non-asymptotic analysis of random matrices*, chapter 5 of Compressed sensing, theory and applications. Y. Eldar and G. Kutyniok. Cambridge University Press, 2012.

[32] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.

[33] K. Wang. Random covariance matrices: universality of local statistics of eigenvalues up to the edge. *RMTA*, 1(1), 2012.